THEY EXPECT RESULTS…

Perspectives on Academia, Industry, and Network Security Monitoring

Christian Kreibich
christian@icir.org
@ckreibich

Your speaker

ICSI

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

lastline

corelight

YES
NYY 5
PITCH 84
TB 1
▲4
MLB.com

Part 1

NEGATIVITY

IN ACADEMIA

INDUSTRY:

This can **work**, let’s use it!

ACADEMIA:

This can **fail** **(I think)** . . .

STRONG REJECT!

Consequence: *group-think*

- "IDS signatures are stupid"
- "Threat intelligence is useless"
- "Network ML is futile"
- "Dynamic analysis doesn't work"
- "Only people over 50 study TCP"

Instead: challenge assumptions

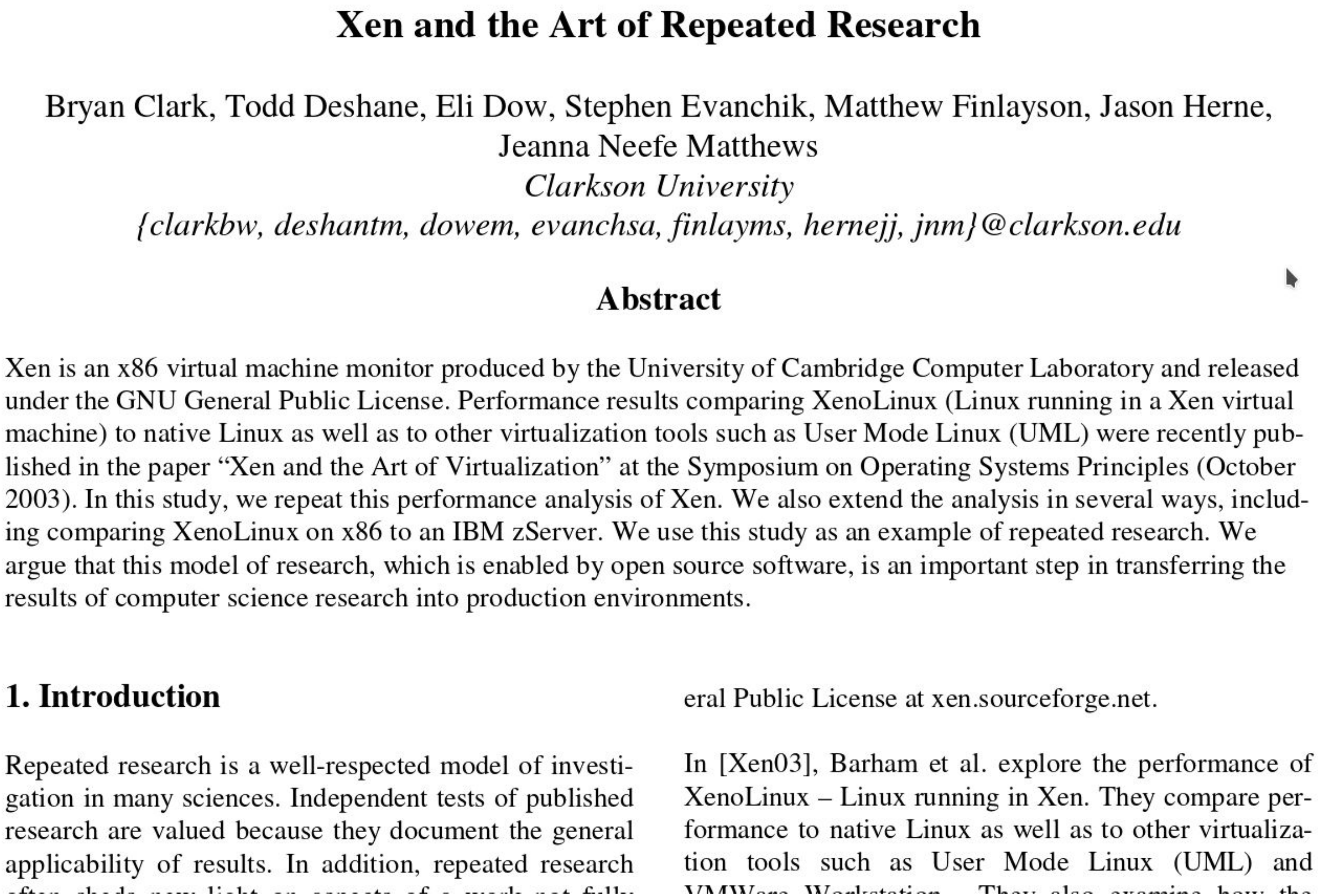
- Study solutions assumed to be “done”
  - nmap wasn’t actually that fast
  - TLS implementations, OMFG
- Prove assumptions wrong in certain settings
  - Signatures work great for protocol detection
- Assume cutting edge becomes status quo
  - Thesis tip anno 2003: leverage pervasive virtualization

Consequence: *culture of negativity*

- Program committees regularly reject publishable work
- Junior researchers mistake negativity for competence, also fear competition
- Negativity is a career killer for both researchers and entrepreneurs

Consequence: *shallow knowledge*

- “This just got published in ACM OAKNIX!!1!”
- Repeated research is highly valuable
  - We are lousy at providing repeatable results (for valid and not so valid reasons)
  - Papers that try to reproduce results are *rare*



Xen and the Art of Repeated Research

Bryan Clark, Todd Deshane, Eli Dow, Stephen Evanchik, Matthew Finlayson, Jason Herne,
Jeanna Neefe Matthews
Clarkson University
{clarkbw, deshantm, dowem, evanchsa, finlayms, hernejj, jnm}@clarkson.edu

Abstract

Xen is an x86 virtual machine monitor produced by the University of Cambridge Computer Laboratory and released under the GNU General Public License. Performance results comparing XenoLinux (Linux running in a Xen virtual machine) to native Linux as well as to other virtualization tools such as User Mode Linux (UML) were recently published in the paper "Xen and the Art of Virtualization" at the Symposium on Operating Systems Principles (October 2003). In this study, we repeat this performance analysis of Xen. We also extend the analysis in several ways, including comparing XenoLinux on x86 to an IBM zServer. We use this study as an example of repeated research. We argue that this model of research, which is enabled by open source software, is an important step in transferring the results of computer science research into production environments.

1. Introduction

Repeated research is a well-respected model of investigation in many sciences. Independent tests of published research are valued because they document the general applicability of results. In addition, repeated research often sheds new light on aspects of a work not fully

eral Public License at xen.sourceforge.net.

In [Xen03], Barham et al. explore the performance of XenoLinux – Linux running in Xen. They compare performance to native Linux as well as to other virtualization tools such as User Mode Linux (UML) and VMWare Workstation. They also examine how the

MANET Simulation Studies: The Incredibles [*]

Stuart Kurkowski
skurkows@mines.edu

Tracy Camp
tcamp@mines.edu

Michael Colagrosso
mcolagro@mines.edu

MCS Department, Colorado School of Mines, Golden, Colorado, USA

Simulation is the research tool of choice for a majority of the mobile ad hoc network (MANET) community. However, while the use of simulation has increased, the credibility of the simulation results has decreased. To determine the state of MANET simulation studies, we surveyed the 2000-2005 proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). From our survey, we found significant shortfalls. We present the results of our survey in this paper. We then summarize common simulation study pitfalls found in our survey. Finally, we discuss the tools available that aid the development of rigorous simulation studies. We offer these results to the community with the hope of improving the credibility of MANET simulation-based studies.

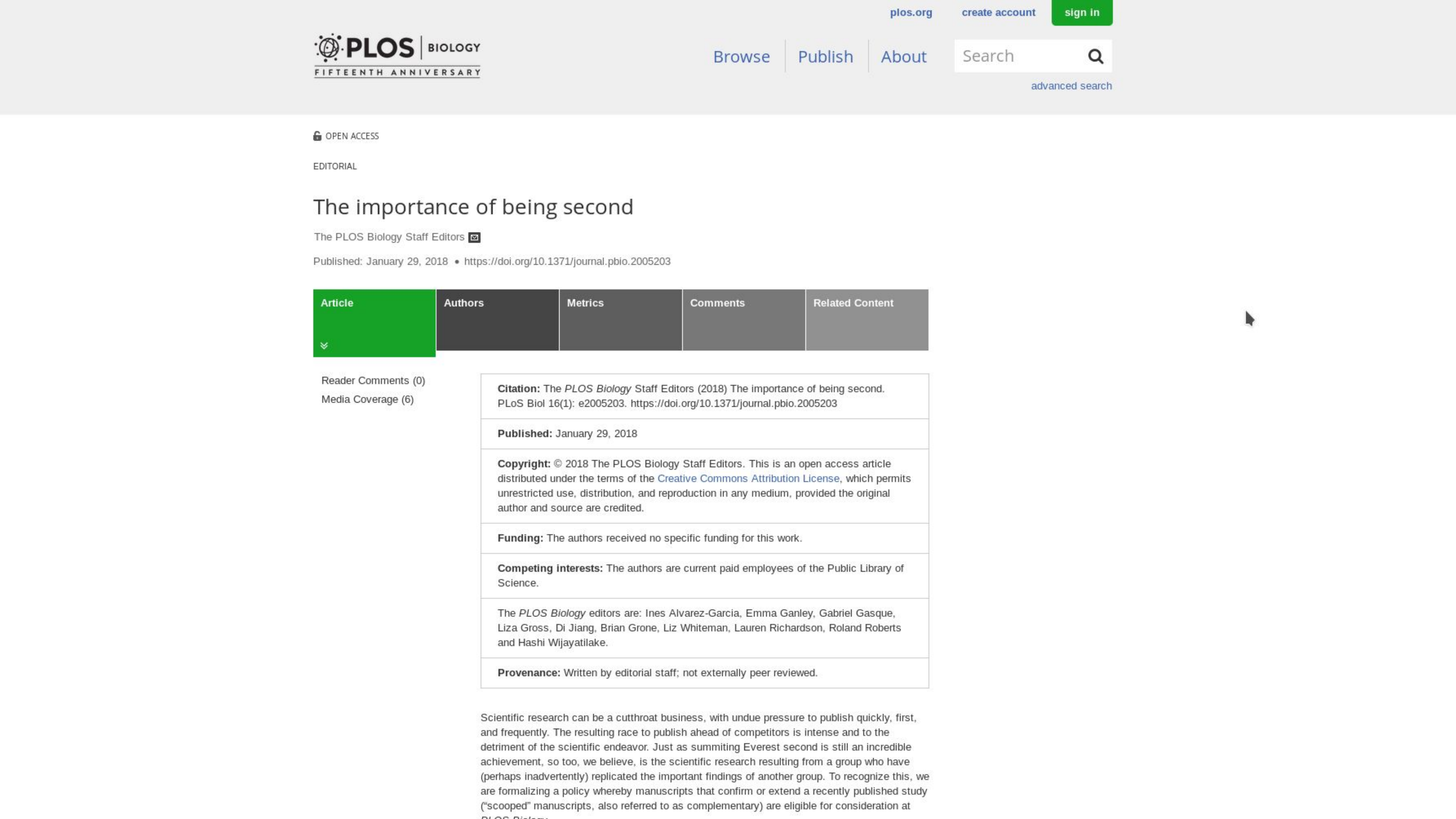
I. Introduction

Mobile Ad Hoc Networks (MANETs) are wireless mobile nodes that cooperatively form a network without infrastructure. Because there is no coordination or configuration prior to setup of a MANET, there are several challenges. These challenges include routing packets in an environment where the topology is changing frequently, wireless communications issues, and resource issues such as limited power and storage.

4. Statistically sound: The execution and analysis of the experiment must be based on mathematical principles.

The remainder of the paper will focus on the current state of MANET simulations, our survey results, common pitfalls to avoid, and tools to aid the researcher in conducting simulation studies. The goal of this paper is to raise awareness on the lack of reliability of MANET simulation-based studies. We present our

Meanwhile in a galaxy far, far away….



OPEN ACCESS

EDITORIAL

The importance of being second

The PLOS Biology Staff Editors

Published: January 29, 2018 • https://doi.org/10.1371/journal.pbio.2005203

Citation: The *PLOS Biology* Staff Editors (2018) The importance of being second. PLoS Biol 16(1): e2005203. https://doi.org/10.1371/journal.pbio.2005203

Published: January 29, 2018

Funding: The authors received no specific funding for this work.

Competing interests: The authors are current paid employees of the Public Library of Science.

The *PLOS Biology* editors are: Ines Alvarez-Garcia, Emma Ganley, Gabriel Gasque, Liza Gross, Di Jiang, Brian Grone, Liz Whiteman, Lauren Richardson, Roland Roberts and Hashi Wijayatilake.

Provenance: Written by editorial staff; not externally peer reviewed.

Scientific research can be a cutthroat business, with undue pressure to publish quickly, first, and frequently. The resulting race to publish ahead of competitors is intense and to the detriment of the scientific endeavor. Just as summiting Everest second is still an incredible achievement, so too, we believe, is the scientific research resulting from a group who have (perhaps inadvertently) replicated the important findings of another group. To recognize this, we are formalizing a policy whereby manuscripts that confirm or extend a recently published study ("scooped" manuscripts, also referred to as complementary) are eligible for consideration at *PLOS Biology*

MÁS AMOR POR FAVOR

Part 2

Industry

vs.

Academia

Industry

- Engineering
- Product management
- Marketing
- Sales
- Customer success
- Finance

ACADEMIA

- Grad student, advisor
- Grad student, advisor
- Grad student, advisor
- Grad student, advisor
- Grad student, advisor
- Grad student, advisor

Industry

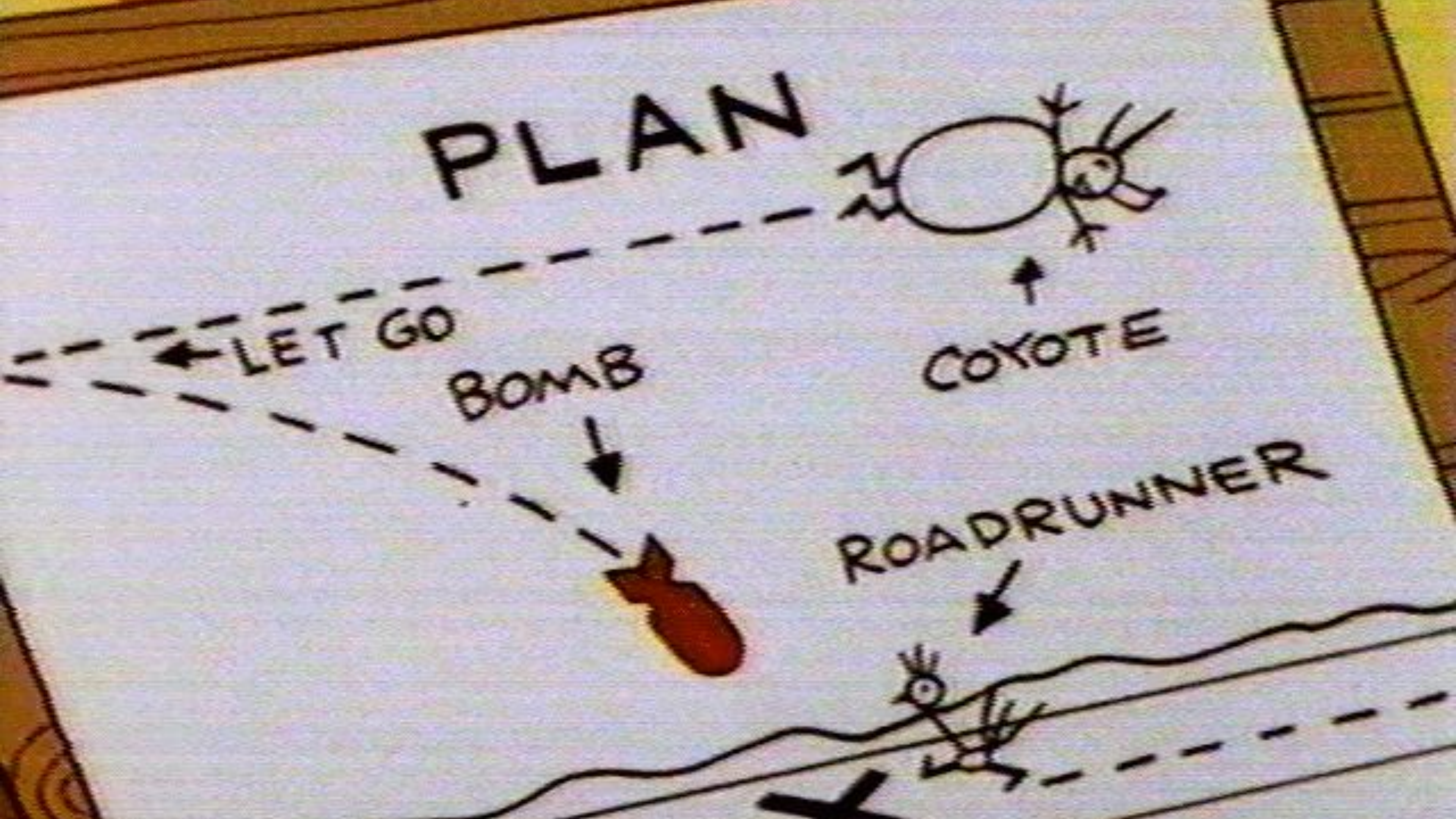
- Engineering
- Product management
- Marketing
- Sales
- Customer success
- Finance

Academia

- Code, experiments
- Project roadmap, collab
- Writing, talks, outreach
- Writing, talks
- Collab, tech support
- Proposals, budgeting

A research group is a small startup, KIND OF

So what if you
work like one?


PLAN
LET GO
BOMB
COYOTE
ROADRUNNER

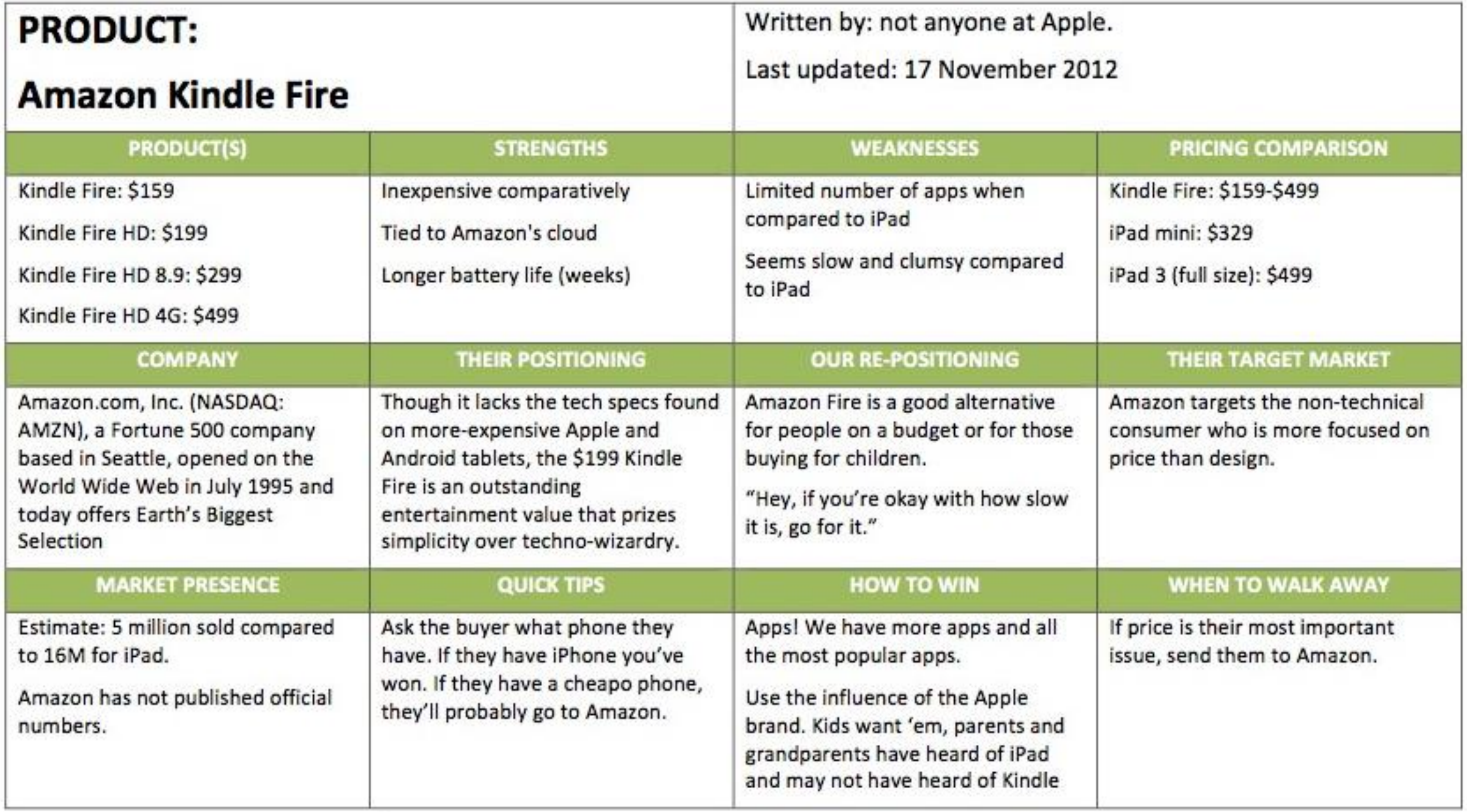
Product

- Do competitive analysis

PROJECT

- Do competitive analysis

“Competitive battle card”



PRODUCT:

Amazon Kindle Fire

Written by: not anyone at Apple.

Last updated: 17 November 2012

PRODUCT(S)	STRENGTHS	WEAKNESSES	PRICING COMPARISON
Kindle Fire: $159 Kindle Fire HD: $199 Kindle Fire HD 8.9: $299 Kindle Fire HD 4G: $499	Inexpensive comparatively Tied to Amazon's cloud Longer battery life (weeks)	Limited number of apps when compared to iPad Seems slow and clumsy compared to iPad	Kindle Fire: $159-$499 iPad mini: $329 iPad 3 (full size): $499
COMPANY	**THEIR POSITIONING**	**OUR RE-POSITIONING**	**THEIR TARGET MARKET**
Amazon.com, Inc. (NASDAQ: AMZN), a Fortune 500 company based in Seattle, opened on the World Wide Web in July 1995 and today offers Earth’s Biggest Selection	Though it lacks the tech specs found on more-expensive Apple and Android tablets, the $199 Kindle Fire is an outstanding entertainment value that prizes simplicity over techno-wizardry.	Amazon Fire is a good alternative for people on a budget or for those buying for children. “Hey, if you’re okay with how slow it is, go for it.”	Amazon targets the non-technical consumer who is more focused on price than design.
MARKET PRESENCE	**QUICK TIPS**	**HOW TO WIN**	**WHEN TO WALK AWAY**
Estimate: 5 million sold compared to 16M for iPad. Amazon has not published official numbers.	Ask the buyer what phone they have. If they have iPhone you’ve won. If they have a cheapo phone, they’ll probably go to Amazon.	Apps! We have more apps and all the most popular apps. Use the influence of the Apple brand. Kids want ‘em, parents and grandparents have heard of iPad and may not have heard of Kindle	If price is their most important issue, send them to Amazon.

Product

- Do competitive analysis
- Study market size
- Create product roadmap
- Plan / build / test
- Release product
- Monitor revenue / success
- Post-mortem

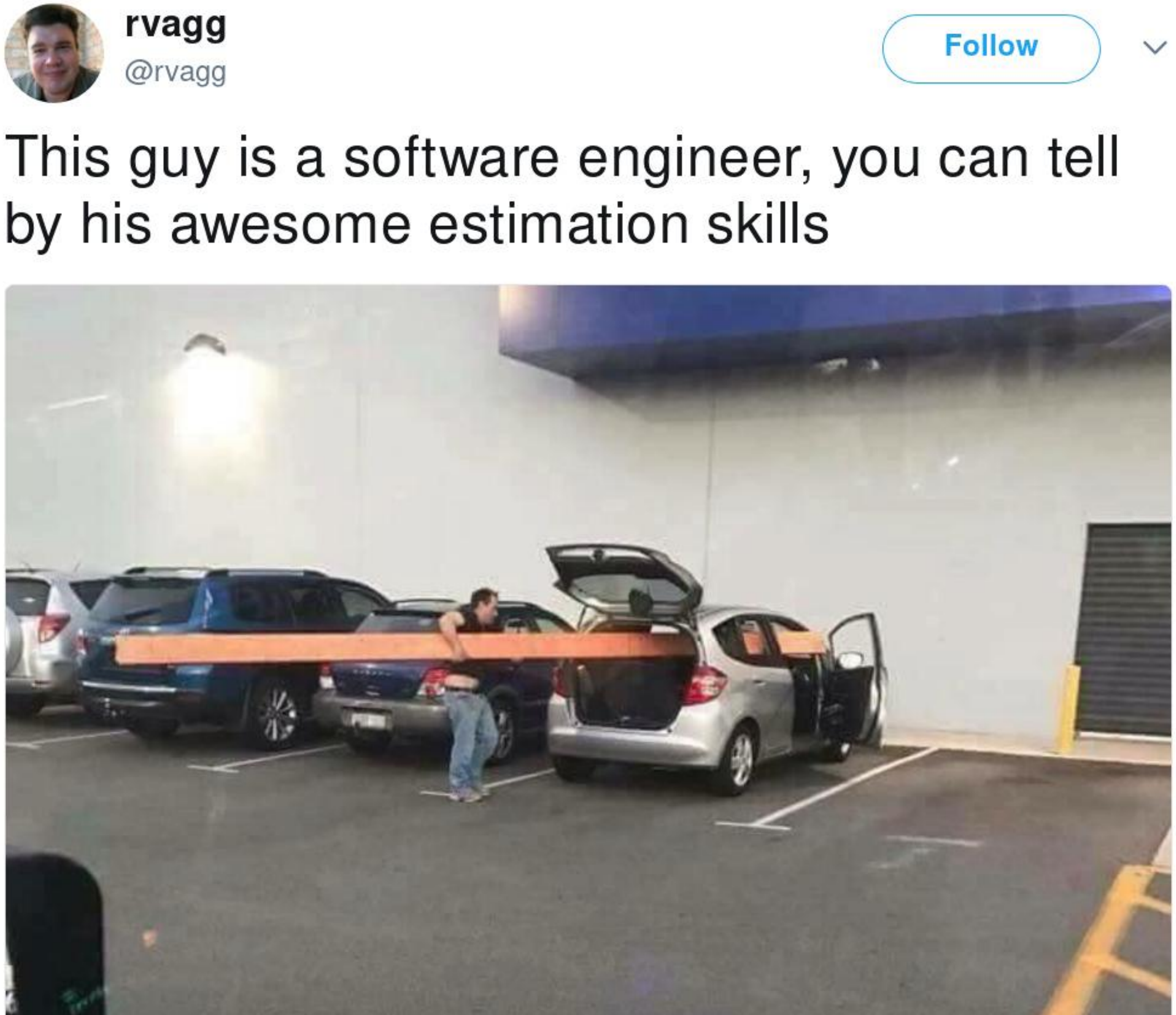
PROJECT

- Do competitive analysis
- Consider venues
- Create paper roadmap
- Hack / experiment / measure
- Submit paper
- Keep submitting paper ...
- Post-mortem

But academia is not industry,

THANKFULLY !

ON



rvagg
@rvagg

This guy is a software engineer, you can tell by his awesome estimation skills

Serendipity

Crazy risk

Freedom

Industry has data

Academia has cycles

Intern early, intern often

… if for research work, check **whether you can publish!**

Industry

- Engineering
- Product management
- Customer success
- Various dedicated roles

ACADEMIA

Grad student, advisor

"The" career in industry does not exist — specialize!

Part 3

Research Suggestions

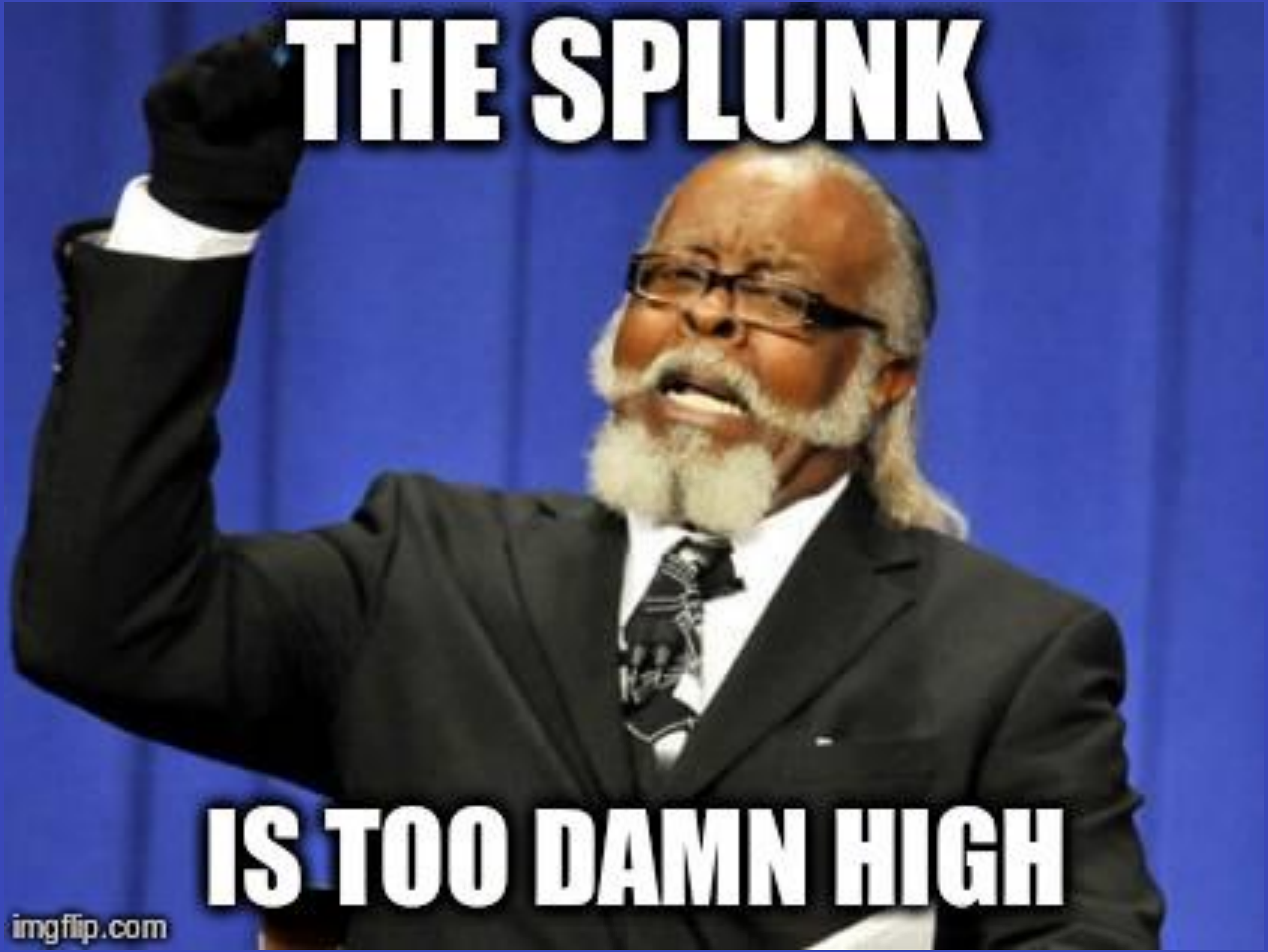
Topic 1: wasteful data storage

Aspect: data compaction/compression in data lakes

Topic 1: wasteful data storage

Aspect: data compaction/compression in data lakes

- ? How to rank, structure, index, control, expire
- ✓ Real-world problem
- ☒ Difficult to access


THE SPLUNK
IS TOO DAMN HIGH
imgflip.com

Topic 1: wasteful data storage

Aspect: efficient storage of network traffic

- ? Structure-aware compression schemes
- ? What would “lossy” mean here?
- ✓ Real-world problem
- ✓ Accessible
- ☒ Perceived as niche problem

Topic 2: corporate networks

- ? Characterize traffic quality, quantity, control
- ? Study locality (physical, data center, on-prem vs cloud)
- ? Revisit assumptions (serverless computing, …)
- ✓ Under-explored & fascinating
- ☒ Difficult to access

Topic 3: middle-box-aware crypto

Large organizations won’t give up DPI

- ? How to manage this better than TLS terminators?
- ? What content do you really need?
- ✓ Immensely relevant
- ❎ Hugely controversial

Topic 4: holistic security

Real-world security is applied risk management

Security infrastructure automation is all the rage

- ? Which defenses work against which threats
- ? What input do these techniques operate on
- ✓ Under-explored
- ☒ Difficult to measure

Additional advice

Keep it simple.

Just imagine you actually need to run that stuff.

If somebody already knows the answer, it’s not research — just go ask![*]

[*] unless they won’t tell you

[*] or they lie :)

New data is better than a new algorithm.

Consider:

- Bro logs
- Datacenter systems work
- Google’s older work on web security
- Research platforms: Netalyzr, Ark, Dasu, Anubis, ...

Stay positive.

Think like a startup, but don’t overdo it.

There’s plenty to do.

THANK YOU

Thanks to Vern Paxson, Robin Sommer, Johanna Amann, Seth Hall, Eric Wolford, Chris Kruegel, Narseo Vallina, Jon Crowcroft, all positivity-encouraging PC chairs I’ve served under, the Corelight product team, many folks who’ve written industry-vs-academia blog posts, the photographers on unsplash.com, Gary Dunaier (aka Thumbs Down Guy), the Wooden Beam Guy, the scientist stock models, and many others...